Determination of Protein Thiol Reduction Potential by Isotope Labeling and Intact Mass Measurement

Determination of protein thiol reduction potential by isotope labeling and intact mass measurement.

Sophie E. Thurlow,[†] David P. Kilgour,[‡] Dominic J. Campopiano,[†] C. Logan Mackay,[†] Pat R. R. Langridge-Smith,[†] David J. Clarke,[†]* Colin J. Campbell.[†]*

EaStCHEM School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh EH9 3FJ, U.K. ;[†] Chemistry and Forensics, Rosalind Franklin Building, Nottingham Trent University, Clifton Campus, Clifton Lane, Nottingham, NG11 8NS, U.K.[‡]

KEYWORDS: Redox potential, protein, mass spectrometry, isotope labeling, thiol, oxidation, redox switch

ABSTRACT: Oxidation/reduction of thiol residues in proteins is an important type of post-translational modification that is implicated in regulating a range of biological processes. The nature of the modification makes it possible to define a quantifiable electrochemical potential, $E^{\oplus}$, for oxidation/reduction that allows cysteine-containing proteins to be ranked based on their propensity to be oxidized. Measuring oxidation of cysteine residues in proteins is difficult using standard electrochemical methods but recently top-down mass-spectrometry has been shown to enable the quantification of $E^{\oplus}$for thiol oxidations. In this paper we demonstrate that mass spectrometry of intact proteins can be used in combination with an isotopic labeling strategy and an automated data analysis algorithm to measure $E^{\oplus}$for the thiols in both *E Coli* Thioredoxin 1 and *Human* Thioredoxin 1. Our methodology relies on accurate mass measurement of proteins using LC-MS analyses and does not necessarily require top-down fragmentation. As well as analyzing homogeneous protein samples, we also demonstrate that our methodology can be used to determine thiol $E^{\oplus}$ measurements in samples which contain mixtures of proteins. Thus the combination of experiential methodology and data analysis regime have the potential to make such measurements in a high-throughput manner and in a manner more accessible to a broad community of protein scientists.

Over the past three decades oxidative thiol modification has emerged as a central mechanism for dynamic post-translational regulation of protein activity and protein signaling pathways. In 1992, Stamler *et al.* investigated the S-nitrosylation of protein cysteine thiols as a reversible covalent modification important in vasodilation.[1] In 2000, Finkel published a review article in which he suggested 'the covalent addition of glutathione to reactive cysteines may be one mechanism for achieving reversible redox-dependent signaling as well as a potential means for identifying relevant redox-dependent signaling molecules.[2] Since then, it has become apparent that many important biological pathways are subject to redox regulation through such oxidation-reduction reactions.[3-5] For example, in the intrinsic apoptotic pathway, the apoptosis regulator protein Bax has been shown to undergo oxidative activation through disulfide formation; and activation of the initiator caspase Casp9 has been shown to occur via intermolecular disulfide formation with apoptotic protease activating factor 1 (APAF1).[6,7] Interestingly, redox regulation seems to be particularly prevalent amongst transcription factors (TF), where cysteine modifications can control function *via* several mechanisms. For example, in the AP-1 family of TFs, Cys oxidation has been shown to induce conformational change, inhibiting DNA-binding.[8] In contrast, Nuclear factor-like 2 (NRF2) activity is controlled by oxidation of its partner protein KEAP1; disulfide formation in KEAP1 results in a reduced affinity for NRF2, allowing translocation of NRF2 into the nucleus and transactivation of target genes.[9] In a recent study of the tumor suppressor TF p53, we demonstrated that oxidation of cysteines within the binding site of the essential zinc ion, results in metal loss, a reduction in protein stability and impaired DNA-binding.[10]As well as modulating protein-protein interactions, there are also many examples of redox regulation of enzyme activity. Of particular note, examples of both kinases and phosphatases have been shown to be regulated by redox chemistry – highlighting how the cellular oxidative environment can potentially influence the phosphoproteome. A clear example of this is found in the members of the PI3K–AKT–mTOR signaling pathway, which is important for regulation of the cell

cycle. The kinase Akt2 is inhibited by oxidation, either through the formation of a disulfide bond or the oxidation of a thiol to a sulfenic acid – both of these modifications are chemically reversible under biological conditions.[11] In addition, the activity of the tumour suppressor phosphatase PTEN, which is the main regulator of the PI3K–AKT–mTOR pathway, can be controlled by both disulfide formation or oxidative modification of its active site cysteine residue.[11]

As a consequence of the central importance of this emerging regulatory mechanism, there is a growing need for chemical and analytical tools for the identification and characterization of redox sensitive cysteines. Here we outline a mass spectrometry based method for the characterisation and quantification of Cys redox modification based on their biochemical standard potential ($E^{\oplus}$) and demonstrate the utility of the procedure using the well-characterized protein thioredoxin. The protein thioredoxin is ubiquitous across the kingdoms and acts as a reductant for a number of proteins, in a variety of biological pathways and it is thought to play an important role in the regulation of apoptosis.[11,12] Its active site contains a CXXC motif with the cysteine residues responsible for its redox activity. As well as the catalytically active residues, human thioredoxin 1 (hTRX1) contains an additional three cysteine residues (Cys61, Cys68 & Cys72). The role of these non-active site residues is less well known; however, Watson *et al*. have determined that Cys61 and Cys68 are capable of forming a disulfide bond.[13]

In biological systems, the propensity for a thiol to undergo oxidative post-translational modification can be quantified in terms of its biochemical standard redox potential $E^{\oplus}$. $E^{\oplus}$ is defined under standard biochemical conditions where all reactants are at unit activity except H^+ which has a concentration of 10^{-7} M (*i.e.* pH = 7). The relationship between $E^{\oplus}$, the relative concentrations of reduced and oxidized proteoforms and the environmental potential (E) is described by the Nernst Equation. For example for a protein with two cysteines that can form an intramolecular disulfide the standard potential can be defined as the potential at which there are equal concentrations of the disulfide and dithiol forms:

$$E = E^{\oplus} + \frac{RT}{nF} ln \frac{[Disulfide]}{[Dithiol]} \qquad \text{Equation 1}$$

From Equation 1 two things are clear:

- If E can be controlled and the ratio of the oxidized and reduced proteoforms measured, $E^{\oplus}$ can be experimentally determined.
- knowing $E^{\oplus}$for a protein is important because it allows one to predict the most thermodynamically stable proteoform in a particular biological environment.

Typically, measuring $E^{\oplus}$ for a protein is not trivial. Electrochemical techniques such as cyclic voltammetry are well suited to metal-centred redox reactions, but tend not to work particularly well for protein thiols as the kinetics of electron transfer are slower.

A variant of Western Blotting has been employed to measure protein $E^{\oplus}$ and an advantage of the redox blotting technique is the ability to probe the redox state of a specific protein from a complex mixture such as cell lysate.[13] However, the technique assumes that an antibody binds with equal affinity to differently alkylated and differently oxidized proteoforms which may not be true for all antibodies.

Recently, we highlighted the potential of using mass spectrometry as an analytical technique for the characterization of redox proteins.[3] MS has the ability to determine the chemical nature of redox modifications based on accurate protein mass measurements. For example, disulfide bond formation is accompanied by the loss of two hydrogen ions (a mass decrease of 2 Da) and glutathionylation involves the addition of GSH to a specific Cys side chain via a disulfide bond (a mass increase of 305 Da). Furthermore, protease digestion and/or tandem MS experiments can be employed to locate the site of specific redox modifications within a protein. Finally, we outlined a differential isotope labeling strategy which could allow the accurate quantification of the ratio of the oxidized and reduced proteoforms of a redox active protein, thus allowing the determination of $E^{\oplus}$for a specific cysteine oxidation event.

This methodology was subsequently adapted by Scotcher et al., who used a differential isotope labeling strategy together with top-down mass spectrometry to determine $E^{\oplus}$for thioredoxin.[14]

The inclusion of top-down fragmentation in the workflow has two notable advantages. Firstly, it provides multiple isotopologue product ions pairs, all of which can be used for quantification of the relative ratios of oxidized and reduced proteoforms, thus this methodology provides multiple replicates in a single top-down spectrum. Secondly, fragmentation of the intact protein prior to mass analysis has the potential to delineate multiple redox active sites within a single protein. However, despite these advantages, the use of top-down fragmentation in the workflow limits the experimental throughput. Top-down fragmentation requires skilled user input and comparatively long data acquisition times. Thus the top-down process is not well suited for automated workflows or hyphenation to liquid chromatography separation.

In principle, for proteins with single redox active sites, the ratio of oxidized and reduced proteoforms can be accurately determined based solely on the ratio of light and heavy isotopologue in the intact protein mass spectrum. Here, we show that this is indeed the case and demonstrate the utility of a differential isotope labeling strategy

coupled with intact mass analysis by characterizing the redox potential of cysteines in human thioredoxin. Analysis of overlapping isotope distributions in mass spectra of isotope labeled intact proteins is a greater challenge than with protein fragments and we developed a bespoke software package (NEMESIS) that greatly simplifies and automates this process. In principle, this workflow is applicable to a range of LC-ESI-MS instrumentation and works with mixtures of proteins. The use of LC-MS for data acquisition and tailored software for data analysis allows the process to be automated; thus opening the door for higher throughput analysis of protein redox modifications.

MATERIALS & METHODS

Materials

Escherichia coli thioredoxin 1 (*E. coli* TRX1) was bought from Sigma Aldrich, resuspended in distilled water at 100 µM, aliquoted and stored at -20 ˚C. All solvents were LC-MS grade unless stated otherwise.

Expression of recombinant wild type and active site mutant human TRX1 in E. coli.

BL21(DE3) competent cells (Agilent Technologies) were transformed with pET16a plasmids containing the hTRX1 gene or the active site mutant hTRX1 (C31,34S) with N-terminal 6His tags provided by the Blom group.[15] 500 ml cultures of 2YT media were used for protein expression. Cells were grown to an optical density of 0.6 at A_{600nm} at 37 ºC before induction with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested after 3 hours. After freezing at -20 °C, the cell pellet was resuspended in chromatography buffer A (20 mM Tris pH 7.4, 0.5 M NaCl, 20 mM imidazole) and sonicated for 15 minutes (30 seconds on, 30 seconds off) to achieve lysis. After removal of the cellular debris, the supernatant was filtered before loading onto a 1ml HisTrap HP column (GE Healthcare).

An AKTA purification system (GE Healthcare) operated by Unicorn was used for affinity chromatography. After protein loading, the column was washed to baseline before an elution gradient from 0 – 100% buffer B (20 mM Tris pH 7.4, 0.5 M NaCl, 1 M imidazole) over 40 column volumes was performed. Protein containing fractions were pooled and loaded onto a HiLoad Superdex 75 (16/60) for size exclusion chromatography and eluted over 1.5 column volumes with 50 mM Tris-HCl, pH 7.0, 150 mM NaCl. Again, protein containing fractions were pooled and concentrated with 5 KDa molecular weight cut off spin filters. The protein was stored at -20 °C.

Redox Titration

Redox buffered solutions consisting of either DL-dithiothreitol (DTT_{RED}) and 4,5-dihydroxy-1,2-dithiane (DTT_{OX}) in 50 mM Tris pH 7.0 or glutathione (GSH) and glutathione disulfide (GSSG) in 100 mM sodium phosphate pH 7.0 were set up as detailed in Tables S1-S4. The Nernst equation was used to calculate the solution potential (considering an $E^{\oplus}$ of -0.330 V vs NHE for the DTT redox pair).[16] The protein (10 µM) was equilibrated for 120 minutes at 25 °C before the reduced cysteine residues were alkylated with 5 mM of the first alkylation reagent - either*N*-ethyl-d_5-maleimide (dNEM) or *N*-ethylmaleimide (NEM). Alkylation was allowed to proceed for 30 minutes. The alkylation reaction was quenched by precipitating the protein with the addition of an equal volume of ice cold 20% (w/v) trichloroacetic acid (TCA). The resulting precipitate was resuspended and reduced with 8 mM TCEP. After reduction was complete, alkylation of the remaining cysteine residues was achieved by addition of 10 mM of the second alkylation reagent (NEM or dNEM) and incubation for 30 minutes.

Each redox condition was performed in triplicate and the mean of the three experimental replicates was used to fit the sigmoidal curve. The standard deviation of the experimental replicates was used as the error for each specific potential and the errors quoted for $E^{\oplus}$ are the errors associated with the fit of the sigmoidal function. For the avoidance of doubt, all potentials reported in this paper are measured relative to the normal hydrogen electrode (NHE).

It is important to note that this labeling strategy is independent of the order of NEM and dNEM labeling. In initial experiments we showed that the use of light or heavy NEM could be interchanged in the procedure without influencing the final result (data not shown).

LC-MS Analysis

High performance liquid chromatography was performed on a Dionex Ultima 3000 (Thermo Scientific) using a reverse phase monolithic PS-DVB (500 µm × 50 mm) analytical column. Buffer A consisted of H_2O & 0.03 % trifluroacetic acid, and Buffer B consisted of 100 % acetonitrile & 0.03 % trifluoroacetic acid. The column was operated at 60 °C with a flow rate of 15 µl/min. Typically, 5 µl injections of 5 uM protein were performed. After protein loading, the column was washed for 5 minutes at 0 % buffer B and 5 minutes at 20 % buffer B before a linear elution gradient from 20 – 100 % buffer B was applied over 15 minutes. The instrument was operated using Chromeleon and for the on-line coupling of LC to MS, the Bruker Daltonics software 'HyStar' was used.

Mass spectrometry was performed on a SolariX FT-ICR mass spectrometer equipped with a 12 Tesla superconducting magnet (Bruker Daltonics). The instrument was operated in the positive ion mode with a capillary voltage of 4 kV. The drying gas was set to 180ºC and 4L/min. The nebuliser gas pressure was 5 bar. Typically ions were accumulated for 200ms-1000ms and data was collected at 1MWord.

Data Analysis

LC-MS data was analyzed in DataAnalysis software (Bruker Daltonics) and deconvoluted mass spectra were obtained using the maximum entropy (MaxEnt) algorithm. These spectra were then analysed using in house NEMESIS software (see Results and Discussion for details).

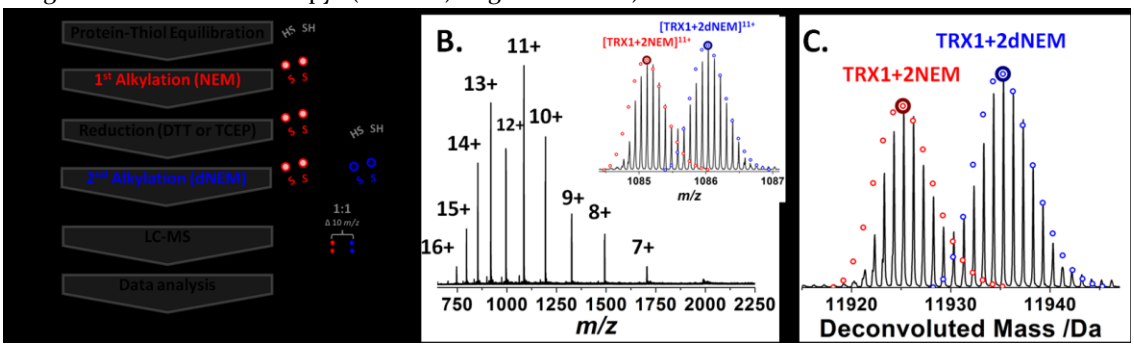

Figure 1: The differential cysteine labeling strategy and typical MS analysis using *E. coli* TRX1 equilibrated at -280 mV as an example. A) An outline of the differential labeling workflow. B) The typical charge state distribution for *E. coli* TRX1, the insert shows the isotope distribution of the 11+ charge state and highlights the two labelled proteoforms of *E. coli* TRX1. C) The maximum entropy deconvoluted mass of the alkylated protein.

RESULTS & DISCUSSION

Differential Alkylation strategy and Quantification of alkylated proteoforms.

To test our approach to measuring $E^{\oplus}$for protein thiols we chose *E. coli* Thioredoxin (*E. coli* TRX1) since it is a well characterized simple protein with only two thiols (Cys32 and Cys35) that are known to form a single intramolecular disulfide bond (see table S5 for protein sequence). By incubating the protein in a DTT buffer of potential -280 mV vs NHE and using the isotopic labeling protocol we can see two distinct populations by intact protein mass spectrometry. LC-MS reveals two similar charge state distributions separated by a mass difference of 10 Da (Fig 1) which corresponds to the difference in mass between thioredoxin labeled with 2 NEM and thioredoxin labeled with 2 dNEM. This observation confirms that, at this specific redox potential, *E. coli* TRX1 exists as two distinct proteoforms in solution and that they can be discriminated using isotopic labeling and intact protein MS. In common with all isotope labeling MS strategies, we assume that the ESI ionization potential of these two species after alkylation is identical and thus ratiometric determination of the MS signal abundance of the two species allows their relative ratio to be accurately determined.

Data Analysis

It is clear for Fig 1B and 1C that the isotope distributions for the two proteoforms of thioredoxin overlap in the mass spectrum. Due to the widening of isotope distributions as molecular weight increases, this phenomenon will become more pronounced in larger intact protein systems. In order to determine the ratio of proteoforms at a given potential we need to separate the overlapping populations of alkylated proteins. To facilitate this in an automated manner suitable for high-throughput analysis we developed "NEM Evolutionary System for Isotopic Solutions" (NEMESIS) which is bespoke software, developed in LabVIEW (National Instruments, Austin, Tx, USA). NEMESIS uses a genetic algorithm to produce an approximately pareto-optimal solution to the relative intensities of the overlapping isotopic distributions recorded in a given spectrum. The genetic algorithm has been described previously[17] and was adapted for use by the development of a new fitness function for this task.

The workflow for NEMESIS starts with loading the deconvoluted mass spectrum in profile form, for the region containing the overlapping isotopic distributions of interest. A simple peak detection algorithm is use to identify the position and normalized intensity of the peaks in this region.[18] The assumption is made that the spectrum is of a resolving power such that the peaks in the distributions are sufficiently resolved so that the peak centroid positions are not limited by resolving power, but not sufficiently resolved such that the isotopologue peaks of the same nominal mass from adjacent distributions can be distinguished by the small mass differences between them. This means that the intensities of overlapping isotopologue peaks from adjacent distributions are taken to sum to give the intensity seen in the mass spectrum for that mass. It is also assumed that the spectrum is not heavily contaminated with other ions.

The protein sequence is entered using single letter abbreviations and NEMESIS identifies the number of cysteine residues present. NEMESIS uses the Mercury algorithm developed by Rockwell, Orden and Smith,[19,20] to calculate the normalized isotopic distribution of the ions produced from this sequence with every possible combination of NEM and dNEM for that protein, assuming that there is one addition of either NEM or dNEM at each cysteine. So, for a system with three cysteines, there would be four possible labelled combinations, where L denotes NEM and H denotes dNEM: [3L,0H], [2L1H], [1L2H] and [0L3H]. Or, more generally, for a protein with *n* cys resi-

dues there will be n+1 possible combinations of light and heavy labelled cys numbering from 0 (all light NEM) to n (all heavy dNEM). The mass of the peaks in each distribution are rounded to nominal masses so that they can be easily aligned with the spectral masses. The isotopic distributions should be generated such that the modelled, synthetic spectra produced match the spectral resolution of the recorded spectrum.

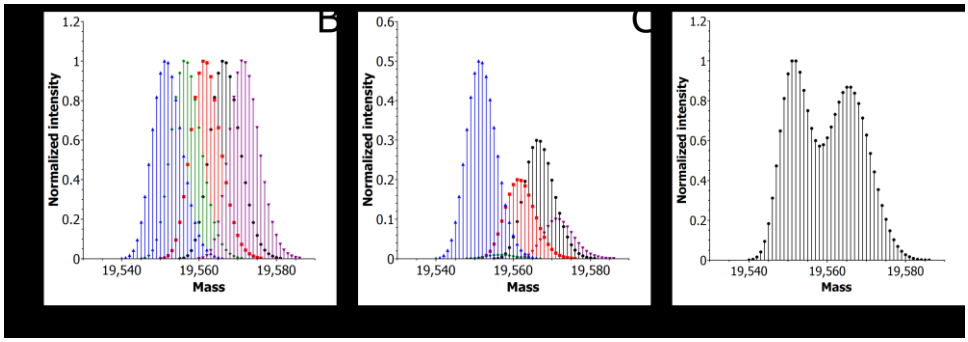

Figure 2: Key steps in the NEMESIS algorithm. A) A synthetic mass spectrum showing the normalized modelled isotopic distributions of the five possible combinations resulting from differential labeling of the four cysteine residues in an example protein comprising 100 Tyr and 4 Cys. B) The effect of scaling the distributions from (A) by the relative proportions 0.5, 0.01, 0.2, 0.3 and 0.1. C) The normalized total synthetic spectrum resulting from summing the distributions in (B).

The complete distribution of peaks in the recorded mass spectrum comes from the sum of the intensities of ions from all combinations at each nominal mass. NEMESIS attempts to find the relative proportions of each combination that, when all the peak intensities are summed appropriately, would give the closest possible approximation of the recorded mass spectrum. An example of this process is shown in Figure 2 for an example protein comprising $Tyr_{100}Cys_4$, where (A) shows the initial modelled (i.e. synthetic) distribution for the 5 possible combinations; (B) shows the combinations scaled to proportions of 0.5, 0.01, 0.2, 0.3 and 0.1 for [4L, 0H], [3L, 1H], [2L, 2H], [1L, 3H] and [0L, 4H] respectively; and (C) shows the total distribution that would result from summing the separate distributions from (B).

A modelled version of the complete peak distribution can be simply defined by the relative proportions of each of the combinations described above, as shown in Figure 2 (C). We can define this total distribution using a simple ordered list just by recording the relative proportion (I) for each combination: [I_0, I_1, ... I_n]. Each different possible distribution can be defined in this way, and in the terminology of genetic algorithms, this is the genetic representation of each solution. So, the genetic representation of the distribution shown in Figure 2 (C) would be [0.5, 0.01, 0.2, 0.3, 0.1].

NEMESIS then tests a series of generations of possible solutions (defined by their genetic representations) to the recorded mass spectrum – the first generation is created randomly (n=50) where each I_n in each member is a randomly generated fraction between 0 and 1. In all subsequent generations of solutions, the members of the new population are generated from the fittest solutions from the previous generation as described previously. The fitness of each member in the population of solutions, in every generation, is estimated simply by the normalized root mean square error between the recorded spectral peak intensities and the sum of the modelled distributions in each case for each nominal mass. Only those masses that correspond to peaks in the modelled, synthetic spectra are considered – not all masses across the complete mass range of the recorded mass spectrum. The fittest members of every generation are preferentially used to generate the next generation in a conventional genetic algorithm approach, using exhaustive cross-over and low probability random mutation (probability p=0.25, mutation range, r=1). The fittest member of each generation is compared to the fittest member of the previous generationand once the fitness ceases to increase by more than 0.05% from generation to generation, the system is deemed optimized and the genetic representation of fittest member of the final generation is output as the solution.

One benefit of genetic algorithm optimization is that the technique is robust to changes in initial parameters. For example, changing any (or indeed all) of the parameters of the genetic algorithm used in NEMESIS by an order of magnitude makes no statistical difference to the final result although it may take more generations to reach the solution.

Redox Titrations and the Determination of $E^{\oplus}$ by Intact Protein Mass Measurement – *E. coli* Thioredoxin.

We then performed the differential labeling protocol on *E. coli* TRX1 incubated in the DTT redox couple at nine defined reduction potentials ranging from -309 to -257 mV. After intact mass analysis, only two protein populations are observed at all potentials tested (Fig 3). The masses of these populations are consistent with intact *E. coli* TRX1 with the two cysteines alkylated with either two dNEM or two NEM labels. It is clear that a smooth transition occurs between the two proteoforms as redox potential increases. Furthermore, the observation that both cysteines in the protein transition from the oxidised to reduced state at the same potential demonstrates that the two residues share the same mid-point potential, indicative of the formation of a disulfide bond between the two residues. This is in agreement with the literature and the catalytic mechanism of the protein.

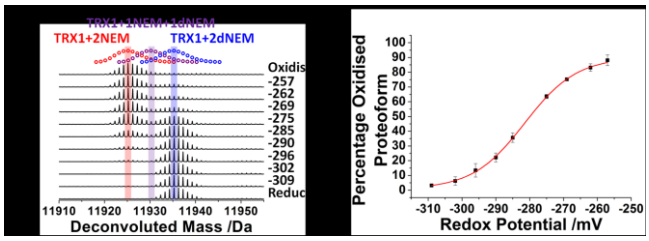

Figure 3: Determination of $E^{\oplus}$ in E. coli TRX1. A) A waterfall plot showing the deconvoluted mass spectrum of E.coli TRX1 at different reduction potential and after differential NEM labeling. Note: dNEM was used in the first alkylation step. B) A plot of the percentage of the oxidised proteoform against the redox potential of the initial equilibration buffer (relative to the standard hydrogen electrode). A sigmoidal trendline was fitted to the data, error bars show 1 standard deviation for n=3.

The ratio of oxidised to reduced *E. coli* TRX1 was then calculated by using NEMESIS to determine the ratio of the two dNEM to two NEM labelled species. By plotting the percentage of the oxidised proteoform of the protein against the redox potential of the DTT solution, it is possible to determine $E^{\oplus}$from a sigmoidal trendline (in this case the point on the trendline at which there are equal amounts of reduced and oxidised proteoforms). From the trendline fitted in 3B, the mid-point potential of the two cysteine residues in *E. coli* TRX1 was found to be -281.5 ± 0.2 mV. This is in good agreement with both established and recent literature publications which have established this potential to lie between -270 mV and -284 mV.[14]

The method described here involves a number of advantages over established techniques used to investigate redox proteins, most notably: the low cost and availability of the reagents, the small sample volumes required, and the automation of data processing.

***Human* Thioredoxin**

In contrast to E. coli TRX, the human form of Thioredoxin 1 (hTRX1) has five cysteine residues (Table S5). hTRX1 contains the characteristic N-terminal CGPC motif (with Cys31 and Cys34) as well as three further C-terminal Cys residues (Cys61, Cys68 and Cys72) Therefore this protein constitutes a more challenging target for analysis using our methodology. By equilibrating hTRX1 in the same DTT buffer system as we used for *E. coli* TRX1 and employing the same dual labeling protocol, we see that there are two proteoforms present over this range of potentials – corresponding to the fully reduced hTRX1 and hTRX1 containing two thiols that oxidise at the same potential (characteristic of disulfide formation). By plotting the ratio of oxidised to reduced proteoforms over this range we can calculate an $E^{\oplus}$ for this process of -276 ± 1 mV (Fig 4).

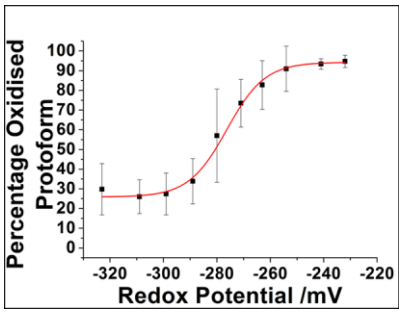

Figure 4: Determination of $E^{\oplus}$ of Cys31 & Cys 34 in hTRX1 by plotting the percentage of the oxidised proteoform as determined using the differential alkylation protocol against the initial equilibration potential. The sigmoidal trendline fitted indicates that the residues have an $E^{\oplus}$ of -276 ± 1 mV. The error bars show one standard deviation from the mean for n = 3 experimental replicates.

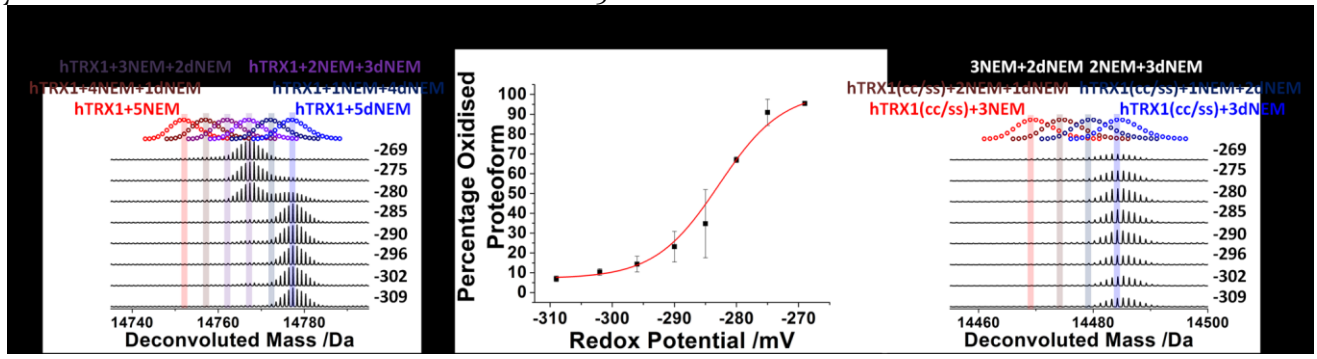

Figure 5: Determination of $E^{\oplus}$ for hTRX1 and hTRX1 (C31,34S). A) A waterfall plot showing the deconvoluted mass spectrum of hTRX between -309 and -269mV reduction potential and after differential NEM labeling. Note: dNEM was used in the first alkylation step. B) A plot of the percentage of the oxidised proteoform (2NEM3dNEM) against the redox potential of the initial equilibration buffer. A sigmoidal trendline was fitted to the data, which indicates that the residues have an $E^{\oplus}$ of -283± 3 mV; error bars show 1 standard deviation for n=3.C) A waterfall plot showing the deconvoluted mass spectrum of hTRX C31,34S mutant between -309 and -269mV reduction potential and after differential NEM labeling, which shows no significant oxidation under these conditions.

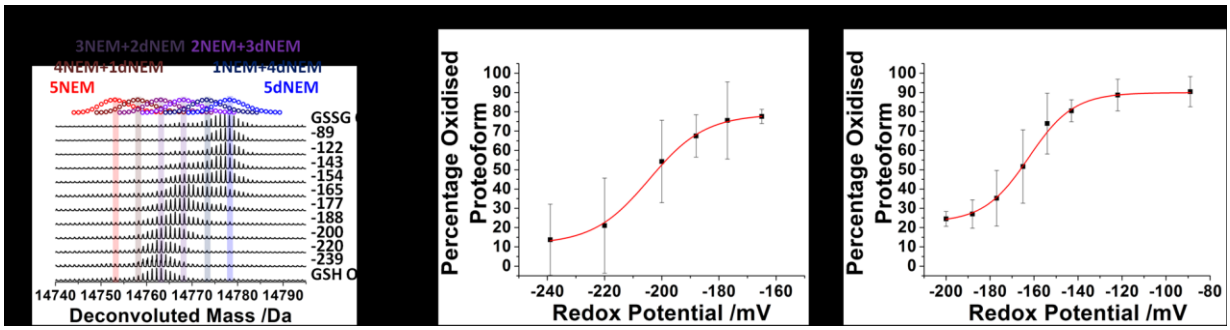

Figure 6: NEMESIS was used to quantify the labeling state of hTRX1 after equilibration in glutathione redox buffers and application of the differential alkylation protocol. (A) A waterfall plot showing the deconvoluted mass spectrum of hTRX1 between -240 and -80 mV and after differential NEM labeling. (B) Determination of the $E^{\oplus}$of Glutathionylation of Cys72. The percentage of the oxidised proteoform hTRX1 with a glutathionylation (labelled as hTRX1+2NEM+3dNEM), relative to the reduced proteoform (hTRX1 with only the active site residues oxidised, hTRX1+3NEM+2dNEM) The sigmoidal line of best fit indicates an $E^{\oplus}$of -204.4 ± 1.4 mV. (C) Determination of $E^{\oplus}$ of the disulfide bond between Cys61 and Cys68. The sigmoidal line of best fit indicates $E^{\oplus}$of -162.4 ±_1.3 mV. (error bars show 1 standard deviation; for n=3 experimental replicates).

This observation is consistent with a recent literature report by Scotcher et al. which demonstrates that the thioredoxin active site cysteines (Cys31 and Cys34) form an intermolecular disulfide bond at -281 mV.[14] .By employing top-down fragmentation we can clearly confirm that the two cysteines which form this disulfide bond are indeed Cys31 and Cys34 (see supporting information Fig S1).

As an alternative approach to confirm that the disulfide forms between the active site cysteine residues, we carried out the same experiment using both the wild type protein (hTRX1) and an active site cysteine to serine mutant hTRX1(C31,34S). The two proteins were mixed and incubated in a series of redox buffers together before MS analysis and data analysis (Fig 5). Although the two species co-elute on the LC column, the mass difference between the two proteins allowed them to be distinguished in the mass spectrometer. Under these experimental conditions we determined the $E^{\oplus}$ of disulfide bond formation in wild-type hTRX1 to be -283± 3 mV (Fig 5A and 5B). It is interesting to note that the presence of a mixture of proteins in the sample had little effect on the $E^{\oplus}$ determine for wild-type hTRX1. Figure 5C shows that over the potential range measured, there is no significant oxidation of the 3 *C*-terminal cysteines in the hTRX1(C31,34S) mutant and confirms that the process observed in Figure 5A is disulfide formation between the two active site cysteines.

These data demonstrate that the intact mass of differentially labelled proteins can be used to determine $E^{\oplus}$for specific cysteine residues. Furthermore we demonstrate the relative ease of adapting our protocol for the determination of $E^{\oplus}$of multiple proteins in the same experiment. Thus highlighting the potential for its future use to investigate electron transfer reactions between proteins.

Since the three *C*-terminal cysteines (Cys61, Cys68, Cys72) are predominantly reduced under the potentials determined by DTT solutions, we used mixtures of glutathione and glutathione disulfide to investigate their redox regulation. This redox couple allows us to set more oxidising reduction potentials. At these more oxidising conditions it was clear that two further redox transitions took place in the C-terminal region of hTRX1. These were identified by top down fragmentation as Glutathionylation of Cys72 and intramolecular disulfide bond formation between Cys61 and Cys68 (see supporting information Fig S2).

Figure 6 shows the deconvoluted mass spectrum for hTRX1 over the potential range -239 mV to -89 mV. Note that for this set of experiments the differential labelling strategy was reversed. The first labeling step, which alkylated the initially reduced cysteines, was performed with NEM; the second labeling step was performed using dNEM.

By using NEMESIS to quantify the differentially labeled protein populations and plotting these against potential we can quantify $E^{\oplus}$ for both the glutationylation and the disulfide formation (Fig 6B and 6C). The $E^{\oplus}$for the glutathionylation is -204.4 mV and that for the disulfide formation is -162.4 mV.

CONCLUSION

We have demonstrated the utility of intact protein mass-spectrometry in the determination of the redox characteristics of a biologically important protein. The technique that we have developed relies jointly on a differential isotope labeling strategy and bespoke software based on a genetic algorithm to determine the ratio of oxidized to reduced proteoforms at potentials applied using soluble redox couples. By plotting the degree of oxidation against potential we can determine $E^{\oplus}$for the protein. We have demonstrated that this protocol can be used to measure $E^{\oplus}$ of proteins in mixtures and to determine $E^{\oplus}$ for multiple redox processes within the same protein. For simple systemBy avoiding fragmentation we have made the process of determining $E^{\oplus}$ for a protein faster and compatible with an LC-MS timescale thus enabling automated, high-throughput analysis. An additional advantage of our workflow is the capability to measure multiple protein redox potentials in the same experiment. The NEMESIS software also greatly simplifies the process of determining $E^{\oplus}$, thus making this type of analysis accessible to a greater range of non-specialist users. NEMESIS is available to academic users on request.

ASSOCIATED CONTENT

Supporting Information

Supplementary information in .doc format.

AUTHOR INFORMATION

Corresponding Author

* E-mail: colin.campbell@ed.ac.uk; dave.clarke@ed.ac.uk

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

CC is grateful to the Leverhulme Trust (Project Grant RPG-2012-680) and EaStCHEM for funding. SET is grateful to the DTC in Cell and Proteomic Technologies and the EPSRC for funding and also to the Blom group for kindly providing the recombinant protein expression plasmids.

REFERENCES

(1) Stamler, J. S.; Simon, D. I.; Osborne, J.; Mullins, M. E.; Jaraki, O.; Michel, T.; Singel, D. J.; Loscalzo, J. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 444-448.

(2) Finkel, T.; Holbrook, N. J. *Nature* **2000**, *408*, 239-247.

(3) Mallikarjun, V.; Clarke, D.; Campbell, C.J. *Free Radic. Biol.* **2012**, *53*, 280-288.

(4) Carroll, K. S.; Paulsen, C. E. *ACS Chem. Biol.* **2010**, *5*, 47-62.

(5) Jones, D. P. *J. Intern. Med.* **2010**, *268*, 432-448.

(6) Zuo, Y.; Xiang, B.; Yang, J.; Sun, X.; Wang, Y.; Cang, H.; Yi, J. *Cell Res.* **2009**, *19*, 449-457.

(7) Nie, C.; Tian, C.; Zhao, L.; Petit, P. X.; Mehrpour, M.; Chen, Q. *J. Biol. Chem.* **2008**, *283*, 15359-15369.

(8) Abate, C.; Patel, L.; Rauscher, F.; Curran, T. *Science.* **1990**, *249*, 1157-1161.

(9) Fourquet, S.; Guerois, R.; Biard, D.; Toledano, M. B. *J. Biol. Chem.* **2010**, *285*, 8463-8471.

(10) Scotcher, J.; Clarke, D. J.; Mackay, C. L.; Hupp, T.; Sadler, P. J.; Langridge-Smith, P. R. R. *Chem. Sci.* **2013,** *4*, 1257-1269.

(11) Truong, T. H.; Carroll, K. S. *Biochemistry* **2012**, *51*, 9954-9965.

(12) Sengupta, R.; Billiar, T. R.; Kagan, V. E.; Stoyanovsky, D. *Biochem. Biophys. Res. Commun.* **2010**, *391*, 1127-1130.

(13) Watson, W. H.; Pohl, J.; Montfort, W. R.; Stuchlik, O.; Reed, M. S.; Powis, G.; Jones, D. P. *J. Biol. Chem.* **2003**, *278*, 33408-33415.

(14) Scotcher, J.; Bythell, B. J.; Marshall, A. G. *Anal. Chem.* **2013**, *19*, 9164-9172.

(15) King, B. C.; Nowakowska, J.; Karsten, C. M.; Köhl, J.; Renström, E.; Blom, A. M. *J. Immunol.* **2012**, *188*, 4103-4112.

(16) T*he Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals*, 12th ed.; Budavari, S.; O'Neal, M.J.; Smith, A.; Heckelman, P. E.; Kinneary, J. F., Eds.; Merck & Co.: Whitehouse Station, NJ, 1996.

(17) Kilgour, D. P. A; Neal, M. J.; Soulby, A. J.; O'Connor, P. B. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 1977-1982.

(18) Kilgour, D. P. A; Wills, R.; Qi, Y.; O'Connor, P. B. *Anal. Chem.* **2013**, *85*, 3903-3911.

(19) Rockwood, A. L.; Van Orden, S. L. *Anal. Chem.* **1996**, *68*, 2027-2030.

(20) Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. *Anal. Chem.* **1995**, *67*, 2699-2704.